Computers & Security 132 (2023) 103302

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

An extended view on measuring tor AS-level adversaries

Gabriel K. Gegenhuber[a,*], Markus Maier[a], Florian Holzbauer[a], Wilfried Mayer[b], Georg Merzdovnik[b], Edgar Weippl[a], Johanna Ullrich[c]

[a] University of Vienna, Research Group Security and Privacy, Kolingasse 14-16, Vienna 1090, Austria
[b] SBA Research, Floragasse 7, Vienna 1040, Austria
[c] Christian Doppler Laboratory for Security and Quality Improvement in the Production System Lifecycle, University of Vienna, Kolingasse 14-16, Vienna 1090, Austria

ARTICLE INFO

Article history:
Received 9 January 2023
Revised 12 May 2023
Accepted 22 May 2023
Available online 25 May 2023

Keywords:
Tor
RIPE atlas
Traceroute measurements
Censorship
Privacy
Anonymity
Routing

ABSTRACT

Tor provides anonymity to millions of users around the globe which has made it a valuable target for malicious actors. As a low-latency anonymity system, it is vulnerable to traffic correlation attacks from strong passive adversaries such as large autonomous systems (ASes). In preliminary work Mayer et al.(2020), we have developed a measurement approach utilizing the RIPE Atlas framework – a network of more than 11,000 probes worldwide – to infer the risk of deanonymization for IPv4 clients in Germany and the US.

In this paper, we apply our methodology to additional scenarios providing a broader picture of the potential for deanonymization in the Tor network. In particular, we (a) repeat our earlier (2020) measurements in 2022 to observe changes over time, (b) adopt our approach for IPv6 to analyze the risk of deanonymization when using this next-generation Internet protocol, and (c) investigate the current situation in Russia, where censorship has been intensified after the beginning of Russia's full-scale invasion of Ukraine. According to our results, Tor provides user anonymity at consistent quality: While individual numbers vary in dependence of client and destination, we were able to identify ASes with the potential to conduct deanonymization attacks. For clients in Germany and the US, the overall picture, however, has not changed since 2020. In addition, the protocols (IPv4 vs. IPv6) do not significantly impact the risk of deanonymization. Russian users are able to securely evade censorship using Tor. Their general risk of deanonymization is, in fact, lower than in the other investigated countries. Beyond, the few ASes with the potential to successfully perform deanonymization are operated by Western companies, further reducing the risk for Russian users.

1. Introduction

Tor is the most notable anonymity network, used by two to three million people every day. A total of 6,500 voluntarily operated Tor relays advertise up to 700 Gbit/s of bandwidth, and provide anonymity by rerouting traffic via three Tor nodes. As a low-latency network, Tor is prone to traffic correlation attacks; thereby, a malicious actor must be able to observe the traffic between the client originating the connection and the first Tor node as well as the traffic between Tor's exit node and the destination. A global passive observer is capable to do so, but this form of an attacker is explicitly excluded from Tor's threat model. Yet, powerful observers exist, potentially threatening the anonymity of Tor users. Their capabilities are, however, not exactly clear. One reason for this is the theoretical assumption that the underlying Internet hierarchy is flat and evenly distributed. This is not the case, as the Internet is shaped in different tiers as well as various entities with different levels of control, e.g., Internet Exchange Points (IXP) with a high level of control and smaller Internet Service Providers (ISPs) with a lower level of control. Also, the Tor network does not utilize the Internet in an evenly distributed manner as the location of Tor relays is depending on various external parameters, e.g., economical (the price of bandwidth) or political (censorship, prosecution) reasons.

Prior work (Edman and Syverson, 2009; Feamster and Dingledine, 2004; Nithyanand et al., 2016) has shown that Tor traffic takes only a limited set of routes on the Internet. These studies,

* Corresponding author.
E-mail addresses: gabriel.gegenhuber@univie.ac.at (G.K. Gegenhuber), markus.maier@univie.ac.at (M. Maier), florian.holzbauer@univie.ac.at (F. Holzbauer), wmayer@sba-research.org (W. Mayer), gmerzdovnik@sba-research.org (G. Merzdovnik), edgar.weippl@univie.ac.at (E. Weippl), johanna.ullrich@univie.ac.at (J. Ullrich).

https://doi.org/10.1016/j.cose.2023.103302

however, rely on BGP updates and route prediction, and claim that measurements – despite being more reliable – would be infeasible due to lacking measurement nodes in the autonomous systems (ASes) that host Tor users, nodes, and destinations. With the introduction of the RIPE Atlas framework (Staff, 2015) – a global measurement network with more than 11,000 probes – this assumption no longer holds. In our preliminary work (Mayer et al., 2020), we developed a measurement methodology utilizing this network to actively probe the Tor network. In more detail, we used the probes to traceroute the Internet paths that are taken by Tor traffic and, based on the collected data, estimated the correlation potential of AS-level adversaries. In comparison to BGP-based approaches of path prediction, active measurements based on traceroute reveal how the packets are actually routed over the Internet. This provides a more realistic risk estimation for Tor users as BGP-based approaches are known to overestimate their risk (Juen et al., 2015).

The paper at hand is an extended version of our preliminary work (Mayer et al., 2020): We apply our methodology to three additional use cases creating an extended view on AS-level adversaries. In particular, we (a) repeat our measurements from 2020 to observe changes in Tor's service quality over time, (b) adopt our approach for IPv6 to analyze the threat of deanonymization when using this next-generation Internet protocol, and (c) investigate the current situation in Russia as censorship has been intensified since the beginning of its full-scale invasion of Ukraine, starting on February 24th, 2022. More specifically, the contributions of this paper are as follows:

- **Updated View on AS Interconnections.** By repeating our measurements from 2020, we investigate whether economical or political factors impacted Tor's service quality. Like in our previous measurements, we identified a few ASes with the potential to successfully deanonymize Tor users; although individual numbers vary over time, the overall picture has remained unchanged. According to our results, Tor provides anonymity at a constant quality to its users in Germany and the US.
- **AS-level Adversaries in IPv6.** We are the very first to conduct active measurements investigating the status quo of IPv6 in the Tor network. Despite the fact that the number of IPv6 Tor relays is smaller than their IPv4 counterparts, we could not identify an increased threat of deanonymization for clients using Tor over IPv6, neither in Germany or the US, nor in Russia.
- **Censorship in Russia.** With Russia's full-scale invasion in Ukraine, Russian state authorities also intensified Internet censorship, i.e., blocking media outlets reporting on the ongoing war. We investigated whether Russian clients evading censorship with Tor are prone to deanonymization, particularly when accessing blocked destinations in their geographic proximity.

The remainder of the paper is organized as follows: Section 2 provides background on Tor, and Section 3 discusses related work. Section 4 explains our measurement methodology. Section 5 provides our measurements' results which are then discussed in Section 6. We outline the limitations of our work in Section 7 and draw our final conclusions in Section 8.

2. Background

Tor was designed by Dingledine et al. (2004) in 2004 and soon became the most popular anonymity system. Tor's protocol specifications are open source and updated on a regular basis (Torproject, 2022b).

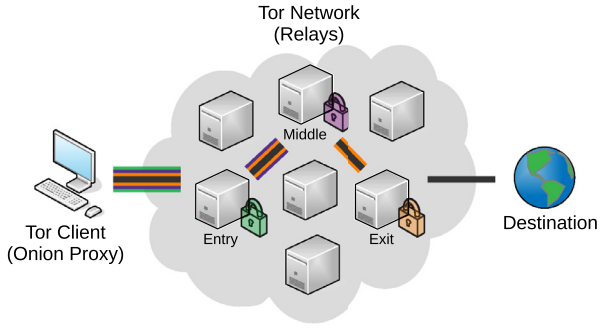


Fig. 1. Tor network. Traffic is relayed via three Tor nodes to hinder correlation of the client and the destination.

Functionality Tor is a low-latency anonymity network based on onion routing. It forms an overlay network of at least three relay nodes that are used to detour user traffic. The entrance to the Tor network is established by the onion proxy, also referred to as the *Tor client*. The proxy handles connections from user applications and is responsible for fetching the initial network information about the Tor network from a set of trusted directory servers. This information is then used to select Tor nodes for relaying. The first relay along a Tor path is called the *guard relay* – it is the only one that knows the client's IP address. The last one on the path is the *exit relay* which is the only one that knows the target IP address. The design of the Tor network is shown in Fig. 1.

Path selection For path selection, the onion proxy relies on information retrieved from the directory servers. The information includes relay flags and bandwidth information about Tor nodes. The exit node is selected first, then the guard relay, and finally the middle relay. The guard- and exit relays are selected randomly; however, the relays are weighted by their bandwidth. The middle relay is selected from the remaining set of nodes. To protect the users and maximize their anonymity, guard- and exit relays are reused according to a strict ruleset (e.g., guard pinning). Additionally, directory servers ensure that only nodes fulfilling certain uptime- and bandwidth requirements are selected as guard nodes. Another requirement for path selection is that the nodes have to belong to different /16 IPv4 prefixes. In reaction to new threat models, these rules are updated frequently, see also Section 3.

Deanonymization of users Tor's design makes it vulnerable to a global passive observer, which monitors all traffic going to and coming from the anonymity network. Such a global observer is explicitly excluded from Tor's threat model; however, powerful observers exist and threaten user anonymity. If an entity is able to monitor both the incoming and outgoing packets of a communication channel, it is able to correlate traffic entering Tor with traffic exiting the network based on timing. Our work precisely focuses on this threat and estimates probabilities of individual ASes appearing in a client's entry- and exit path.

IPv6 support Currently, Tor relays are either operated IPv4 only or Dualstack (i.e., providing an IPv4 and an IPv6 address). This way, Tor allows IPv6 traffic to enter and exit the network. Thereby, Tor relays can also act as bridges between IPv4 and IPv6. It should be noted that connecting from or to an IPv6 address reduces the set of possible relay candidates on the respective connection endpoint.

3. Related work

Feamster and Dingledine (2004) provided the first analysis of location diversity in the Tor network for independently operated ASes based on BGP routing tables. They analyzed the probability of an entry path to the network and an exit path from the net-

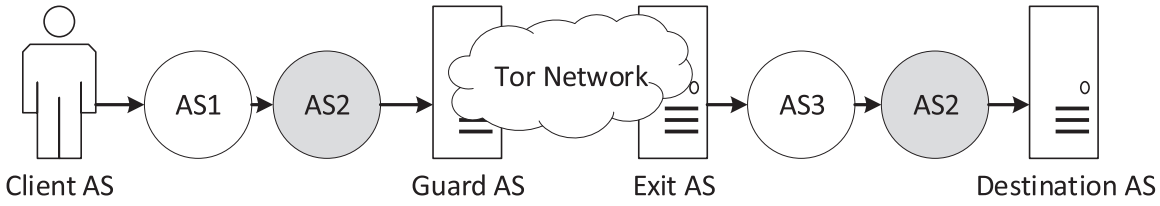


Fig. 2. Threat model. AS2 appears on the Tor entry path, between the client and the guard relay, and on the exit path, between the exit relay and the destination, and is thus in a position to perform traffic correlation deanonymizing the client.

work crossing through the same AS. Their analysis showed that previous methods of choosing paths/nodes based on IP prefixes are not sufficient to guarantee a diverse set of ASes, since there was a 10% to 30% chance, that both the entry and exit path to the mix network crossed the same AS. A refinement of this approach by Edman and Syverson (2009) showed that the previous study had even underestimated the potential threat. A study of Tor security properties against traffic correlation attacks was presented by Johnson et al. (2013). Their results showed that, depending on location, a user's chance of compromise can be at 95% within three months of monitoring against a single AS. One mitigation they proposed is to carefully select which entry and exit nodes to use. Wacek et al. (2013) built a graph of the Tor network to capture the networks' AS boundaries. Using this graph they provided an evaluation of a set of proposed relay selection methods and quantified their respective anonymity properties. Their results showed that bandwidth is an important property for the performance of such algorithms, and should not be neglected.

The importance of location diversity in the Tor network has been shown by several attacks proposed in recent years. Vanbever et al. (2014) provided a study of the capabilities of AS-level adversaries. Sun et al. (2015) described a set of advanced routing attacks on Tor, named *Raptor*. They also described the feasibility of asymmetric AS-level attacks by observing not only data traffic from the exit relay to the server but also TCP acknowledgment traffic on other routes which increases the capabilities of AS-level adversaries. Including the reverse path, they found 31.7% of the Tor circuits to be vulnerable in their measurements. However, paths had different probabilities to be selected by a client, and the actual number was likely to be lower. In 2016, Nithyanand et al. (2016) also used data on the Internet's topology (Giotsas et al., 2014) in a combination with AS-topology simulations (Gill et al., 2012) to estimate the threat posed by adversaries to Tor users. While previous attempts at the correlation of traffic (Hopper et al., 2010; Mittal et al., 2011) had very limited performance or required a large amount of captured traffic or time, *DeepCorr* (Nasr et al., 2018), developed by Nasr et al. greatly improved the feasibility of such attacks. By leveraging emerging learning mechanisms they managed to achieve drastically higher performance compared to existing state-of-the-art systems.

To mitigate the threat of AS-level adversaries that are able to correlate traffic and thereby monitor Tor users, various kinds of protection mechanisms have been proposed (Alsabah and Goldberg, 2016). Nithyanand et al. proposed *Astoria* (Nithyanand et al., 2016), an AS-aware Tor client. While similar in functionality to *LASTor* (Akhoondi et al., 2012), it provided improved protection with concern to threat models and attacker capabilities. Sun et al. (2017) presented a measurement study on the security of Tor against BGP hijacking attacks and presented a new relay selection mechanism to mitigate such attacks on Tor. In contrast to previous approaches, *DeNASA* from Barton and Wright (2016) provided a mechanism for AS-aware path selection independently of the destination. Additionally, they proposed another system for the creation of efficient and anonymous Tor circuits (Barton et al., 2018). Hanley et al. (2019) proposed an extension to the work presented by Sun et al. (2017) to increase the provided privacy and anonymity guarantees. Wan et al. (2019) showed that several attacks against a set of the proposed protections (Counter-RAPTOR, DeNASA, and LASTor) were still possible, but they also proposed simple solutions, which allowed to mitigate the threat posed by their developed methods. Rochet et al. (2020) introduced client-location-aware path selection (CLAPS) to overcome the pitfalls detected in previous path selection solutions (Counter-Raptor, DeNASA). They proved that based on the path selection of the earlier approaches the client's location can be revealed only after a few connections. Eaton et al. (2022) further enhanced the receiver side anonymity of Tor by introducing Private Information Retrieval (PIR) to hide which information is retrieved from the Hidden Service directory servers (HSDirs). Next to security, recent related work also focused on improving the Tor core. Jansen and Johnson (2021) estimated that the actual bandwidth of the Tor network could be much higher. They suggested a new measurement system for bandwidth calculation of Tor nodes. The authors found that with the current system the bandwidth self-measurements resulting in the *observed bandwidth* are rather imprecise.

4. Methodology

In the following section, we describe our method to measure strong AS-level observers, which are in a good position to conduct correlation attacks. As an overlay network, Tor depends on the underlying structure of the Internet. While often a flat hierarchy is assumed, it is clear that this is not the case. We can model the structure of the Internet by looking at autonomous systems identified by a unique AS number (ASN). One AS can be seen as an administrative entity that is responsible for a defined routing policy. Some AS are large and include a lot of Tor users, destinations, or relays, others do not contain users and destinations but are used for routing Tor traffic through the Internet and others are not important for Tor routing at all. Thus, some entities can observe more traffic than others.

With our measurement approach, we find a way to quantify which entities are in a stronger position. Figure 2 illustrates the basic idea of a standard traffic correlation attack, where one adversary (AS2) is placed on the incoming route to Tor as well as on the outgoing route to the destination. Sun et al. (2015) showed that it is also possible to correlate reverse-path traffic that may be routed differently. Other work already quantified strong adversaries with the help of BGP route updates. In contrast, we develop a method that utilizes the RIPE Atlas framework to actively acquire routing

Table 1
Tor relay statistics. While the number of relays increased, they are now spread among fewer ASes. The total Tor bandwidth increased by 66% over the past two years.

(a) Relays				(b) Diff. ASes			(c) Bandwidth (GBit/s)		
	2020	2022	Diff	2020	2022	Diff	2020	2022	Diff
All	6509	6559	*+1%*	1104	981	**-11%**	418	694	*+66%*
Exit	1000	1597	*+60%*	275	222	**-19%**	113	181	*+60%*
Guard	2415	2272	**-6%**	470	469	**-0%**	255	368	*+44%*

Table 2
IPv6 support statistics. As of Sept. 2022, 45% of the Tor relays support IPv6, while the exit bandwidth is 71% of the IPv4's one.

(a) Relays				(b) Diff. ASes			(c) Bandwidth (GBit/s)		
	All	IPv6	Share	All	IPv6	Share	All	IPv6	Share
All	6559	2924	45%	981	375	38%	694	342	49%
Exit	1597	1083	68%	222	94	42%	181	128	71%
Guard	2272	951	42%	469	175	37%	368	152	41%

information as this allows to study how packets actually travel the Internet.

The paper at hand extends our preliminary work; therefore, we apply our measurement approach to three additional use cases to gain a broader view of the potential of deanonymization in the Tor network. (a) We repeat the measurements and compare the state of 2020 with the current state (September 2022). (b) As IPv6 support at Tor relays has improved over the recent years (Torproject, 2021), we adapt our methodology to additionally acquire routing information for IPv6. (c) Lastly, we investigate a practical case study, Russia's full-scale invasion of Ukraine, and analyze AS-level packet routing by simulating access to websites that are blocked by Russian state authorities.

4.1. Relay AS diversity

As shown in Table 1, the Tor network currently (September 17th, 2022) consists of 6,559 relays. Only relays with the *Guard* flag (stable and reliable relays after a ramp-up phase (Dingledine, 2013)), are used as entry relays. Only relays configured to allow exiting traffic are potential exit relays in a Tor circuit. Because of the more stringent requirements, the number of guard and exit relays (with guard/exit probability > 0) is smaller than 6559. This also affects the AS diversity, which is the number of different ASes these relays are placed in.

Table 1 compares the metrics of Tor relay nodes for our two measurement snapshots in 2020 and 2022. Overall, the network has grown in terms of size and offered bandwidth. However, it has become more centralized, as for example, the AS diversity at exit relays has decreased by nearly 20%. Although the number of guard relays dropped by 6%, the number of bridges – providing an alternative and more anonymous entry to the network – nearly doubled (1,350 vs. 2,450) in the respective time period. Since a Tor relay node can – additionally to its IPv4 address – also offer an IPv6 address, we give an overview of the current IPv6 support in Table 2. With IPv6, the AS diversity drops by more than 60% making it an interesting target for our study.

Tor relays are chosen based on their flags and consensus weight. In Fig. 3, we show the AS diversity relation to guard and exit probability. We see that a small number of ASes have a large share of (a) guard and (b) exit probability. For IPv4, only five ASes control more than 50% of exit probability and 43 ASes have more than 90%. We also see that six ASes have a summarized guard probability of more than 50% and 131 have more than 90%. During our measurements in 2020, half of the exit probability was controlled by eight ASes and only four ASes dominated half of the guard probability. Therefore, the accumulated exit probabilities among top ASes has become even more centralized, while the guard probabilities are now slightly more diversified. For IPv6 the centralization is even worse, as only three ASes control more than 50% of guard resp. exit probability. Summarizing, Tor relays are distributed in almost 1000 ASes, the majority of entry and exit routing endpoints are however placed in a few ASes only.

Location diversity provides a similar picture: Two countries (Germany and the US) account for more than 47% of the relays. While the top five countries are still the same as in 2020, we noticed that Russia has lost a majority of its relays and has dropped from the sixth to the 18th rank (from 297 down to just 65 relays).

4.2. The RIPE Atlas framework

The RIPE Atlas framework is a highly distributed measurement network consisting of more than 11,000 available probes, deployed in over 3,600 different ASes. Regarding IPv6, it offers more than 5,000 vantage points (i.e., probes) in over 1,600 different ASes.

The measurement platform allows us to execute various low-level commands, e.g., ping or traceroute, on these probes and further processes the results. We will utilize this to execute traceroute commands from RIPE Atlas probes that are deployed in the same ASes as Tor guard- or exit relays, as well as clients and popular destinations.

Figure 3 also shows the cumulated guard- and exit probability for ASes that contain RIPE Atlas probes. From 222 ASes that contain exit relays, only 98 also contain a probe (837 relays out of 1597). Still, that makes approx. 43% of the total exit probability (35% with only 12 ASes). This differs from the cumulated guard probability. From 469 ASes that contain 2,272 relays, 249 ASes (with 1,723 relays) also include a RIPE Atlas probe, which represents guard relays with a sum of 80% guard probability (60% with 15 ASes). Especially for exit relays, these numbers could be drastically increased if only a few, exit-focused ASes would also host RIPE Atlas probes. Table 3 identifies ASes, that are currently not hosting any RIPE probes. By adding only five probes we could measure ASes with 81% exit probability in total and ten probes would reach 88% probability in total.

Figure 4 shows a bubble graph of current exit relays sorted by AS. The top five ASes that are not covered by RIPE Atlas (cf. Table 3) are marked in red (numbers 1–5). `AS208323 APPLIEDPRIVACY` which is represented by a green bubble (6) was missing RIPE Atlas coverage in 2020, but now hosts a RIPE

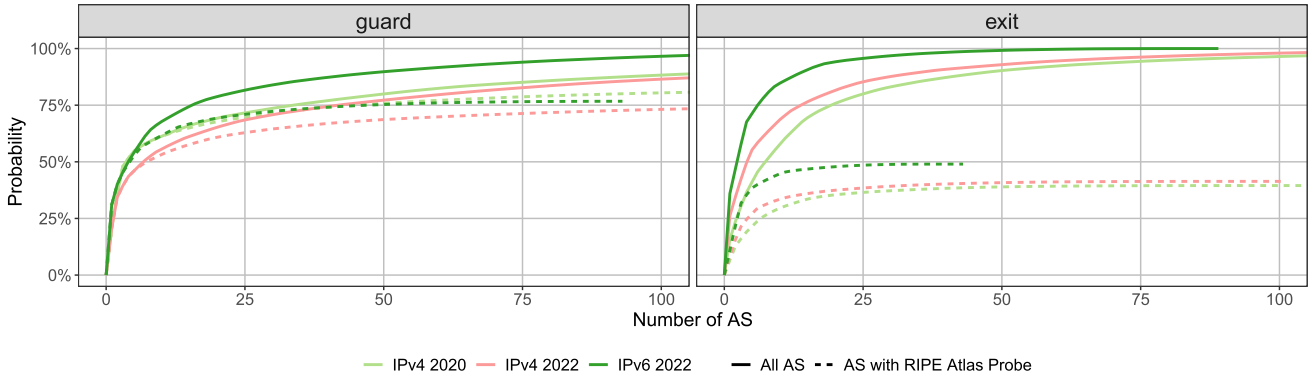


Fig. 3. Accumulated percentage of (a) guard, and (b) exit probability with the number of ASes. While RIPE probes cover 75% of guard probability, they cover less than 50% of exit probability.

Table 3
AS hosting Tor relays but no RIPE Atlas probe. Adding a single probe to AS60729 would increase the accumulated exit probability by 22.1%.

	AS	Name	Relays	Gbit/s	P_{exit}	P_{guard}
Exit	60729	ZWIEBELFR.	225	39.2	0.221	0.002
	205100	F3NETZE	32	11.9	0.084	0.000
	200651	FlokiNET	48	5.95	0.030	0.000
	62744	QUINTEX	100	6.77	0.026	0.000
	4224	CALYX	29	5.36	0.023	0.001
Guard	201814	SKYTECH	81	13.62	0.023	0.018
	46844	SHARKTECH	36	7.17	0.000	0.014
	19437	SS-ASH	10	2.51	0.000	0.006
	200303	LUMASERV	20	2.54	0.000	0.006
	264617	PANAGLOBAL	5	1.83	0.000	0.005

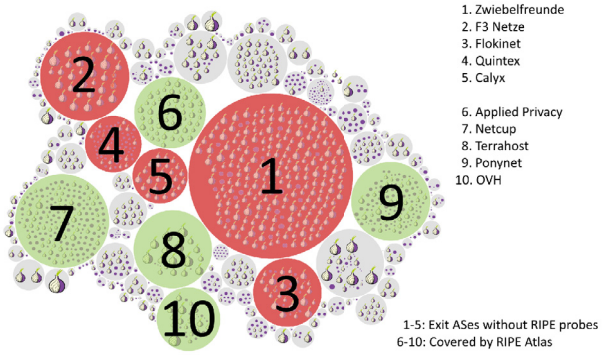


Fig. 4. RIPE Atlas probe coverage of ten large exit relay ASes. Adding a few probes to the exit relay ASes in dark red color (numbers 1–5) could significantly increase the coverage (Metrics, 2022).

probe. Other exemplary exit ASes that are covered by RIPE Atlas are also marked in green (numbers 7–10). Compared to 2020, the overall RIPE Atlas coverage of ASes that contain guard- and exit relays has not changed much (exit: 41 to 43%; guard: 83 to 80%).

4.3. *Active* traceroute *probing with RIPE Atlas*

As illustrated in Fig. 5 we perform *traceroute* measurements to identify routes taken for four different directions: (1) all client ASes to all guard ASes, (2) exit ASes with probes installed to the destination ASes, (3) destination ASes to all exit ASes, and (4) guard ASes with probes installed to the client ASes. With these measurements, we do not cover all possible routes since not all ASes have probes installed. For IPv4, depending on the direction, we measure at step (1) 100%, (2) ~43%, (3) 100%, and (4) ~80% in terms of route probability. For IPv6, we cover at step (1) 100%, (2) ~52%, (3) 100%, and (4) ~85% in terms of route probability.

In detail, this process works as follows:

1. Create the following sets:
   i. AS_{client} ... ASes of the clients (as chosen for the individual scenario, see Section 4.4)
   ii. AS_{guard} ... all ASes with guard relays
   iii. $AS_{guard \cap probe}$ ... all ASes with guard relays and RIPE Atlas probes
   iv. AS_{exit} ... all ASes with exit relays
   v. $AS_{exit \cap probe}$ ... all ASes with exit relays and RIPE Atlas probes
   vi. $AS_{destination}$ ... ASes of the destinations (as chosen for the individual scenario, see Section 4.4)
2. Generate ICMP traceroute measurement definitions for the following directions:
   (1) $AS_{client} \xrightarrow{traceroute} AS_{guard}$
   (2) $AS_{exit \cap probe} \xrightarrow{traceroute} AS_{destination}$
   (3) $AS_{destination} \xrightarrow{traceroute} AS_{exit}$
   (4) $AS_{guard \cap probe} \xrightarrow{traceroute} AS_{client}$
3. Execute the *traceroute* with the RIPE Atlas measurement API (''protocol'': ''ICMP'', ''response_timeout'': 20000, ''packets'': 1).
4. Process all results and look up the corresponding AS from the *ip2asn* database.
5. For every *traceroute*, mark all included ASes with the probability of that path being chosen, i.e., the corresponding guard/ exit probability.
6. Combine the values for directions 1 and 4 for the entry side, and 2 and 3 for the exit side, s.t., if an AS appears on either the forward or the reverse path it is assigned with the probability of that path being chosen. For multiple destinations, all traceroutes are combined.
7. Point out the top ASes, that appear on entry and exit side by looking at $P_{guard} \cap P_{exit}$.

4.4. *Origin and destination AS*

The goal of this work is to investigate multiple Tor scenarios for their proneness to deanonymization attacks. Therefore, we (a) repeat the measurements of our preliminary work in 2020 to observe changes over time, (b) adopt our approach for IPv6 to analyze the threat when using Tor over the next-generation Internet protocol, and (c) extend our measurements to investigate the current situation in Russia as censorship has intensified after its full-scale in-

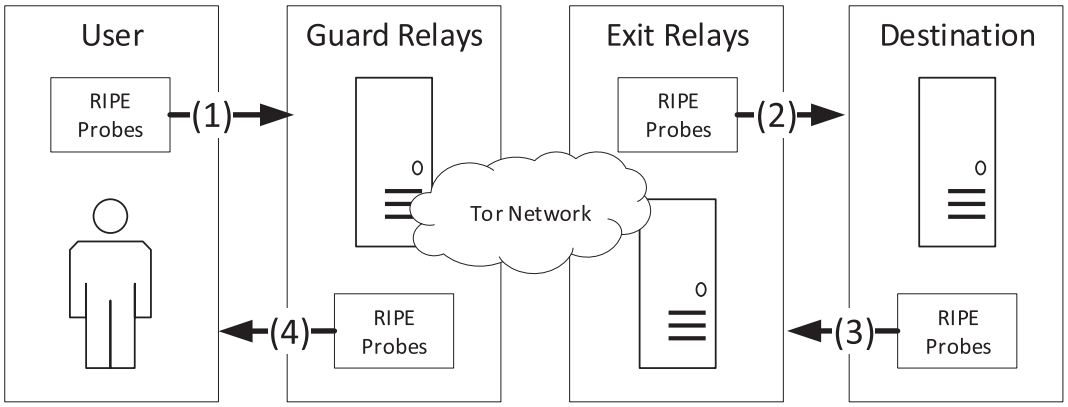


Fig. 5. RIPE Atlas *traceroute* scans. The forward path is covered by D_1, from client to guard relay ASes, and D_2, from exit relay ASes with probes to the destination AS. The reverse path is covered by D_3, from destinations to exit relay ASes with ripe probes, and D_4, from guard relay ASes with probes back to the client.

vasion of Ukraine, starting on February 24th, 2022. The following paragraphs describe how we derived the ASes for our client and destination data sets.

IPv4 measurements As mentioned in Section 4.1 Germany and the US account for nearly half of all Tor nodes. In our 2020 measurements, we have therefore chosen the ten ASes in Germany and the US containing most RIPE Atlas probes – an indicator of the AS's popularity in the respective country – for the client set C. For destinations, we derive the ASes from the Tranco (Pochat et al., 2019) top sites list. In particular, we take the 100 top-ranked domains, resolve the domain, and retrieve the corresponding ASes. We include only those ASes with deployed RIPE Atlas probe(s) in our destination set D. For our *traceroute* measurements, we select one RIPE Atlas probe for each AS in the client and destination set.

For the repetition of our measurements in 2022, we slightly adapted the approach in the following manner: In addition to the ASes inferred according to the just described procedure, we also included ASes that have been investigated in the first iteration (i.e., $C_{ipv4} = C_{2022} \cup C_{2020}$ and $D_{ipv4} = D_{2022} \cup D_{2020}$). For some cases, we were not able to gather updated results for ASes that have been measured in 2020. For example, `AS3356 LEVEL 3` was included in our destination set in 2020, but was not measured in 2022 as it does not host a RIPE Atlas probe anymore. Similarly, the 2020 client set contains historic ASes that do not exist nowadays (e.g., two consumer-grade ASes `AS6830` and `AS31334` were merged to `AS3209` – which is already present in our client set).

IPv6 measurements For clients, we again select ten ASes in Germany and the US with the most RIPE Atlas probes offering IPv6 support. For destinations, we increased the number of included domains from the Tranco list from 100 to 250 due to overall low IPv6 support. For comparison, we also included all ASes supporting IPv6 in the IPv4 datasets and vice versa.

Websites blocked by Russia Russian ISPs had started to block Tor in December 2021 (Xynou and Filastò, 2022b), i.e., three months before the beginning of Russia's full-scale invasion of Ukraine on February 24th 2022. Afterwards, Russia introduced even more rigorous censoring, blocking access to social media and independent news outlets (Migliano and Woodhams, 2022). Many of the blocked destinations are hosted either in Russia or Ukraine and report on the ongoing war. Circumvention of censorship is among the main goals of Tor, making Russia's full-scale invasion of Ukraine an interesting case study. Thus, we investigate whether users from Russian client ASes could be deanonymized when accessing these censored destinations in their geographical proximity.

For the client set, we again determine the ASes with the most RIPE Atlas probes in the respective country, i.e., Russia. For the destination set, we use a public list of websites blocked by Russian state authorities (Xynou and Filastò, 2022a) and rank the domains by popularity using the Tranco list. Then, we resolve these domain names and filter for ASes in Russia or Ukraine. Finally, we match our results with the RIPE Atlas deployment which determines the destination set for this measurement. As there were only two AS candidates supporting IPv6 within this data set, we refrained from a distinct IPv6 measurement in this particular case.

Summary of data set In total, we have eight data sets representing client ASes: 2020 IPv4 Germany, 2020 IPv4 US, 2022 IPv4 Germany, 2022 IPv4 US, 2022 IPv4 Russia, 2022 IPv6 Germany, 2022 IPv6 US, 2022 IPv6 Russia. Please note that there is no 2020 data set for Russia as the country was not included in our previous measurements. Beyond, we have four data sets representing destination ASes: 2020 IPv4 Tranco, 2022 IPv4 Tranco, 2022 IPv6 Tranco, 2022 IPv4 Blocked Websites. A detailed list of the ASes that are included in the data set is found in the Appendix A.

Our approach allows to measure Tor's entry and exit paths independently of each other, and to combine the results only in a successive processing step. Thus, it is sufficient to measure each of the data sets only once.

4.5. Data sources

To facilitate reproducibility and encourage openness, all used data files are publicly available at the project website.[1] In particular, our work relies on following data sources:

1. The Tor consensus, that contains all Tor relays with their IP addresses (IPv4, IPv6), associated flags (particularly "Guard" and "Exit"), advertised bandwidth, and guard and exit probability. We collect this information via the Tor network status protocol *onionoo*.[2]
2. Statistical data about the *RIPE Atlas probes*.[3]We use different data (e.g., id, number and AS of the probes) to find all probes connected to the same ASes as guard and exit relays.
3. Freely accessible *ip2asn*[4] databases to match IP addresses with the corresponding AS number.
4. Active RIPE Atlas *traceroute* results.[5]

[1] Project website: https://www.github.com/sbaresearch/ripe-tor
[2] onionoo: https://www.metrics.torproject.org/onionoo.html
[3] probes: https://www.atlas.ripe.net/probes/
[4] ip2asn: https://www.iptoasn.com/
[5] measurements: https://www.atlas.ripe.net/measurements

Table 4
Entry path and exit path probabilities for a single client and a single destination. HETZNER appears on the entry path due to its high number of entry relays as well as on the exit path due to hosting the destination.

	AS	Name	C	P	P_{relays}	P_{routes}	R
entry	1764	NEXTLAYER	○	1.00	–	1.00	468
	3356	LEVEL3	●	0.330	0.002	0.328	59
	24940	HETZNER	○	0.224	0.224	0.000	2
	16276	OVH	◐	0.131	0.130	0.000	2
	174	COGENT	●	0.123	0.002	0.121	110
exit	24940	HETZNER	○	1.00	-	1.00	220
	60729	ZWIEBELFR.	◐	0.221	0.221	–	1
	25291	INTERDOTL.	●	0.221	–	0.221	1
	47147	AS-ANX	●	0.162	–	0.162	2
	6939	HURRICANE	●	0.130	0.002	0.128	28

○ Client or Destination AS. ◐ Guard or Exit AS. ● Transit AS.

Table 5
AS with the potential for traffic correlation. For all measurements, HETZNER has the potential to deanonymize the client due to appearing on the entry and exit path.

		AS	Name	C	P_{guard}	P_{exit}	$P_{\&}$
2020	v4	24940	HETZNER	○	0.202	0.988	0.199
		1200	AMS-IX1	●	0.180	0.068	0.012
		16276	OVH	●	0.152	0.065	0.010
2022	v4	24940	HETZNER	○	0.224	1.00	0.224
	v6	24940	HETZNER	○	0.350	0.998	0.350
		6939	HURRICANE	●	0.087	0.393	0.034
		47147	AS-ANX ANE	●	0.107	0.223	0.024
		197540	NETCUP-AS	●	0.107	0.139	0.015

○ Client or Destination AS. ● Transit AS.

5. Evaluation

In this section, we discuss the results of our measurements: For readability, we first illustrate our approach in an exemplary measurement including a single client and destination AS only (see Section 5.1). Then, we focus on the full measurement discussing the ASes residing on Tor's entry paths (see Section 5.2), and those on the exit paths (see Section 5.3). Finally, we combine these results to infer ASes appearing on Tor's entry and exit path with high probability as they have the potential to perform traffic correlation deanonymizing Tor users (see Section 5.4).

5.1. Exemplary measurement: single client and destination

As an illustration of the capabilities of our methodology, we evaluate the results of measurements with a single fixed client AS and a fixed destination AS. herefore, we choose the AS of our research center $C = \{AS1764\}$ as client, and the AS of one mirror of the torproject.org website $D = \{AS24940\}$ as destination. In these ASes, we selected RIPE Atlas probes and scheduled 1,194 traceroutes, as defined in Section 4.3; out of which 1,177 (98.6%) were executed successfully (D1: 563/563, D2: 104/109, D3: 240/240, D4: 270/282).

Table 4 shows the results for IPv4; the ASes are grouped depending on whether they reside on a path towards a guard relay, or on a path from an exit relay. As expected, the client resp. destination AS (HETZNER, NEXTLAYER) is found in all traceroutes. Beyond, the ASes HETZNER, OVH and ZWIEBELFREUNDE appear in the tables; serving a high share of guard resp. exit bandwidth in the Tor network, the respective route is likely to be chosen as a Tor path (P_{relays}). However, we want to focus on intermediate ASes, that are different from those hosting relays as well as client/destination and appear on many routes. We identified LEVEL3, COGENT, and HURRICANE to be in a powerful position.

Eventually, we filter for intermediate ASes that have a probability of 1% or higher to appear on both sides, and only a single AS remains, namely HETZNER. With $P_{guard} = 0.224$ and $P_{exit} = 1.000$, it has a chance of $P = .224$ to deanonymize Tor traffic from our research center to torproject.org.

Table 5 provides an overview of the ASes with a probability of 1% or higher to appear on both sides for the three measurements 2020 IPv4, 2022 IPv4, and 2022 IPv6. A comparison of the IPv4 measurements reveals that the number of such ASes decreased; however, the probability of HETZNER increased by 2.5 percentage points. This means that the AS has now an even higher chance of deanonymization due to its increased guard probability. For IPv6-based traffic, this number is even higher. Because the set of possible guard relays decreases in IPv6, the guard probability of HETZNER increases once again by more than 10 percentage points. Beyond, there are three transit ASes in IPv6 with a $P_{\&}$ of up to 3.4%.

The case of HETZNER is particularly interesting as its chance of deanonymization arises from a distinct combination: On the one hand, it is the destination of our measurements; on the other hand, it hosts a high share of guard bandwidth and is thus more likely to be pinned as a guard node. This raises the question of whether path selection should include the destination AS to prevent such scenarios.

5.2. Tor entry: ASes between clients and guard relays

In the following paragraphs, we investigate the chance of ASes to be on a route to/from a guard relay and the chosen client ASes in Germany, the US, and Russia. For a total of 20 intermediate ASes, Fig. 6 shows their entry path probabilities as inferred from our measurements. The 20 AS were chosen according to the following rules: For every country, we select the 15 most likely intermediary ASes. We then show all intermediary ASes that occur for more than five clients in every measured country in our graph. For graphs that correspond to measurements in 2022, we also include ASes that were selected at the previous measurement period. Each data point represents one specific client AS. For a (transit) AS that is present in our measured routes to Tor guard relays, a data point in the graph denotes the summarized probabilities of all routes, i.e., the probability that this AS can trace packets from the client to the Tor network. On the right side of each row, we show the total number of data points. To visualize the range of the single data points, we draw a line between the minimum and maximum values. The figure allows a comparison among our three measurements (2020 IPv4, 2022 IPv4, and 2022 IPv6) as well as among the chosen countries.

In essence, the overall picture for IPv4 was confirmed, and the ASes with high entry path probability in 2022 remain the same as in 2020. For example, any client AS uses – though with varying probabilities – paths including AS174 COGENT, AS1299 TWELVE99, AS3356 LEVEL3. Yet, certain changes were observed: First, AS1200 AMS-IX1, AS12876 ONLINE S.A.S., and AS35807 SKYNET-SPB-AS appeared in the 2020 measurements, but not in the latest of 2022. In return, previously unknown ASes were seen (AS44530 HOPUS HOPUS, AS47147 AS-ANX ANEXIA). Second, AS2914 TT-COMMUNICATIONS-2 was frequently observed for US-based client ASes in 2020; nowadays, its probability is roughly comparable to the Germany-based client ASes.

Comparing IPv4 and IPv6, AS6939 HURRICANE is found more often on paths from US-based client ASes; in return, AS174 COGENT, AS1299 TWELVE99 and AS3257 GTT-BACKBONE are traversed less often. Beyond this fact, Tor paths of IPv4 and IPv6 appear to be highly similar, particularly for client ASes in Germany and Russia.

Local differences Most clients taking routes through high probability transit ASes are from the US: For example, AS6939

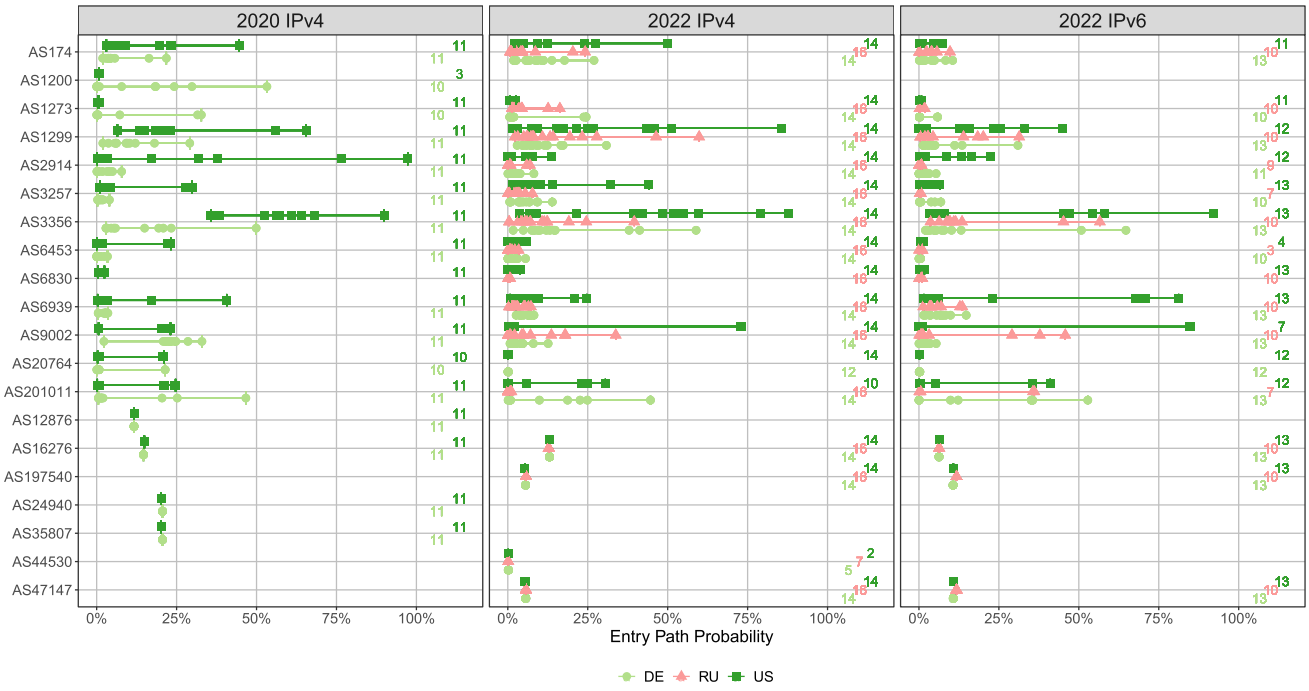


Fig. 6. Entry path probability describing the chance of an AS to appear between the client AS of three different countries and the guard relay. Each data point represents a client AS. The number on the right is the total number of data points.

HURRICANE is particularly dominant for IPv6 in the US, but plays only a minor role for German and Russian ASes. AS9002 RETN-NET plays a strong role in the US but has lower probability in Russia and Germany. This might indicate that routing in the US is more centralized than in the other countries. Beyond, it appears that ASes that are frequently found on paths from German client ASes, are also often seen on paths from Russian ASes; this might be a consequence of their geographic proximity.

While the sole presence of an AS on a path to/from a guard relay is not sufficient to conduct traffic correlation, it might however be sufficient to identify clients – and successively their users – connecting to the Tor network. Thus, the discussed results, covering Tor's entry side, also provide insights on which ASes are capable to detect clients using Tor.

5.3. Tor exit: ASes between exit relays and destinations

In the following paragraphs, we investigate the chance of ASes to be on a route between an exit relay and the chosen destinations for two distinct destinations sets: the Tranco List representing the most popular domains and those officially blocked by the Russian state-authority Roskomnadzor.

Tranco list For a total of 14 intermediary ASes, Fig. 7 shows the probability for the destination ASes that have been inferred from the Tranco list. For every destination we select all ASes that have a maximum probability of more than 20%. To filter for significant ASes we remove rows with a median value of less than 1% or less than five data points. For graphs that correspond to measurements in 2022, we also include ASes that were selected during the previous measurement period. Each data point represents a specific destination AS, and the figure allows a comparison among our three measurements (2020 IPv4, 2022 IPv4, 2022 IPv6). ASes that were selected because they are hosting a substantial amount of exit relays (e.g., AS60729 ZWIEBELFREUNDE) are marked with an asterisk.

For the AS that have already been seen in the 2020 measurements, we see a similar picture in 2022, and only minor changes are apparent: AS6461 ZAYO is barely seen anymore, and AS1200 AMS-IX1 is gone. The latter has also been identified for the entry side. Beyond, we found five new ASes with a considerable chance of being along the path. Comparing IPv4 and IPv6, we see that the maximum probabilities are typically lower for IPv6 for most ASes. Conversely, AS6939 HURRICANE has better chances to be on the path towards the destination, i.e., this AS appears to be a dominant player in the IPv6 Internet.

Destinations blocked by Russia Figure 8 shows the respective probability for the destination ASes that are blocked by the Russian state. As these websites have been predominantly blocked since the start of Russia's full-scale invasion of Ukraine, we do not have any data from 2020. We refrained from measuring IPv6 as only two of the candidate ASes were IPv6-ready. The ASes that are found towards these destinations are also found towards the Tranco List, with a single exception: AS3223 VOXILITY, an Internet infrastructure provider based in UK.

5.4. Potential ASes for traffic correlation

As a final step, we combine the results from Tor's entry paths, between client and guard relays, and exit paths, between exit relays and destinations. We calculate the probability that an intermediate AS is residing on both paths as the latter is the prerequisite to conduct a successful correlation attack deanonymizing the client.

Tranco list Our results are depicted in Fig. 9, providing the respective probability for the three measurements 2020 IPv4, 2022 IPv4, and 2022 IPv6, as well as the three investigated countries Germany, the US, and Russia. Each data point in a graph represents a transit AS that has both entry and exit path probability higher than 0%. For the entry path, we show data points for all relevant clients. For the exit path, we use the maximum probability

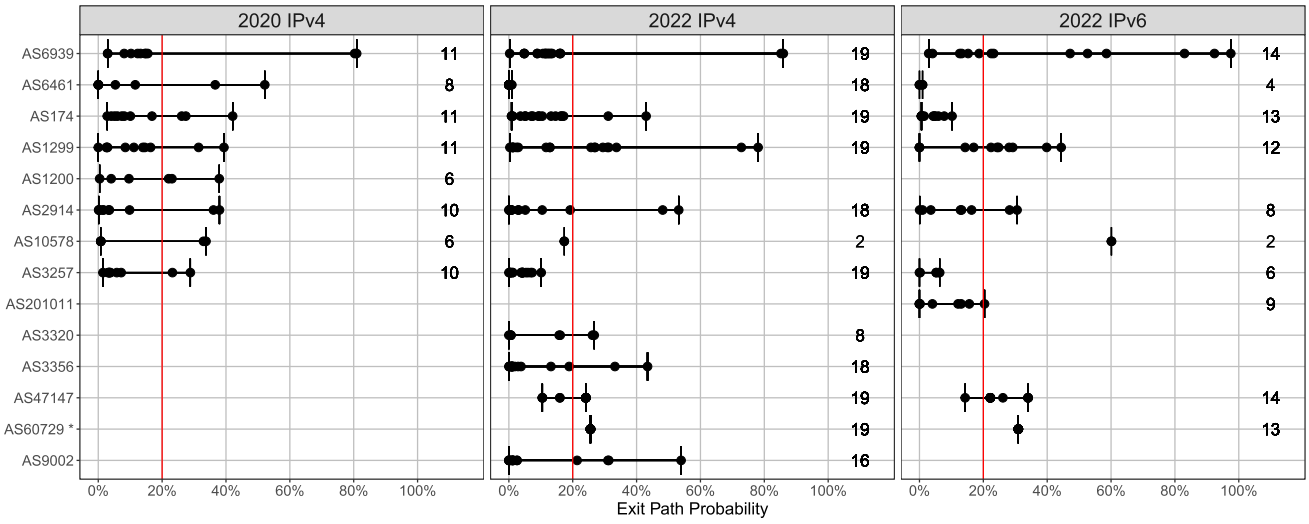


Fig. 7. Exit path probability describing the chance of an AS to appear between the exit relay and the Tranco list destinations. Each data point represents a destination AS. The number on the right is the total number of data points.

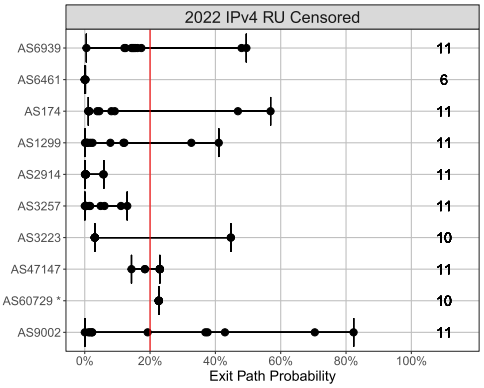


Fig. 8. Exit path probability for the ASes hosting websites that are blocked by Russia. The depicted ASes are all operated by Western companies (i.e., US, SE, UK, AT, DE).

of all measured destination ASes, which represents the worst-case scenario – i.e., an attacker has the best chance to correlate traffic when this target is visited by the Tor user.

In summary, AS24940 HETZNER is strong in all scenarios: First, it serves destinations and is thus likely to be on the exit side. Second, it hosts a high share of guard bandwidth, and is thus likely to serve as a guard relay, eventually appearing on the entry side. In combination, this leads to a high chance of being capable to correlate Tor traffic. As an exception, the measurement on the bottom left (2022 IPv6 DE) shows a reduced exit probability for AS24940 HETZNER. In this case, the selected measurement probe for scheduling traceroutes from AS24940 HETZNER to the Tor network – corresponding to (3) in Fig. 5 – ran into a timeout and did not yield any results. Although we also measure the same routes in the opposite direction – i.e., from the Tor network to AS24940 HETZNER, corresponding to (2) in Fig. 5 – this only covered about 50% of AS probability, due to the lack of RIPE Atlas availability in exit relay ASes (cf. Fig. 3).

Since 2020, the exit probability of AS3356 LEVEL3 has decreased substantially. With this, its overall chance for a successful attack decreased, both for German-based and US-based clients. Yet, this AS has still been considered relevant due to having a good probability to be found on the entry side, particularly in the US. In return, AS1299 TWELVE99 has increased its exit probability in this time span.

For IPv6, we see high chances for US-based traffic to be correlated: AS6939 HURRICANE stands out. It has high exit probability and also respectable entry probability at several client ASes. Beyond, AS3356 LEVEL3 is noteworthy because it has an excellent entry probability for specific client ASes. For both protocols, it appears that there are less chances to correlate traffic originating from Russian-based client ASes.

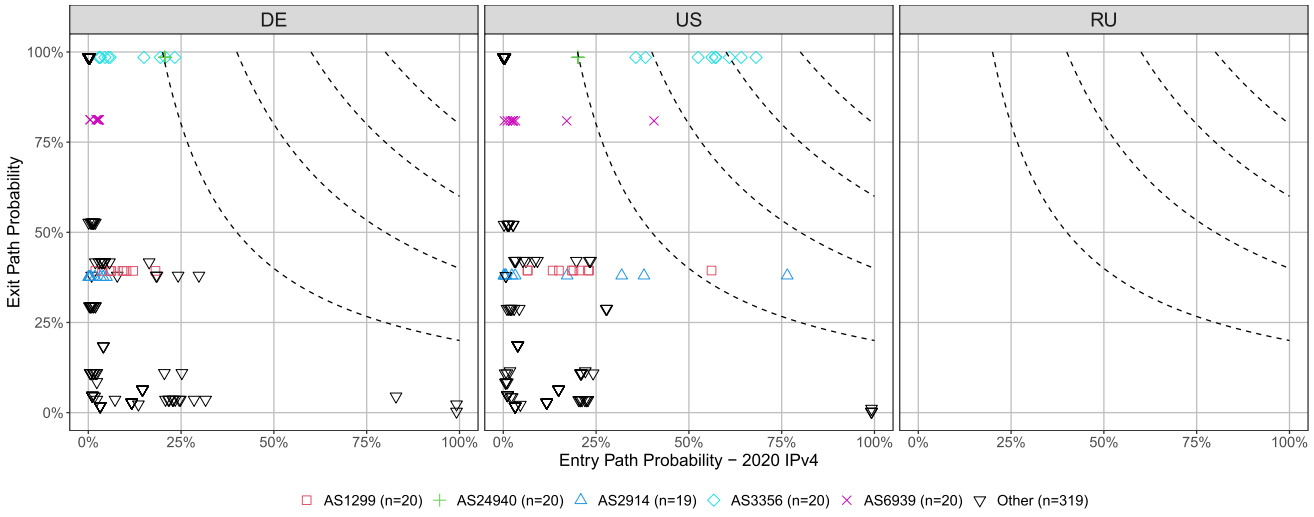
Destinations blocked by Russia The combined probabilities for the ASes hosting websites that are blocked by the Russian state are presented in Fig. 10. Again, the contour lines highlight data points at 20%, 40%, 60% and 80% combined probability. In this case, we see again that there are lower chances to correlate Russian-originating traffic than those from other countries. Although the client ASes (Russia) are within regional proximity to our destination ASes (Russia, Ukraine), the relevant transit ASes do not change much from our previous results. As an outlier, AS20764 RASCOM appears with an exit path probability of 52.3%. It is a consequence of the client AS simultaneously being the transit of the destination and appearing for a single client (AS20764 itself) only.

Consequently, we assume that a regional attacker (e.g., a nation-state) is not able to match entry- and exit packets of local Tor clients.

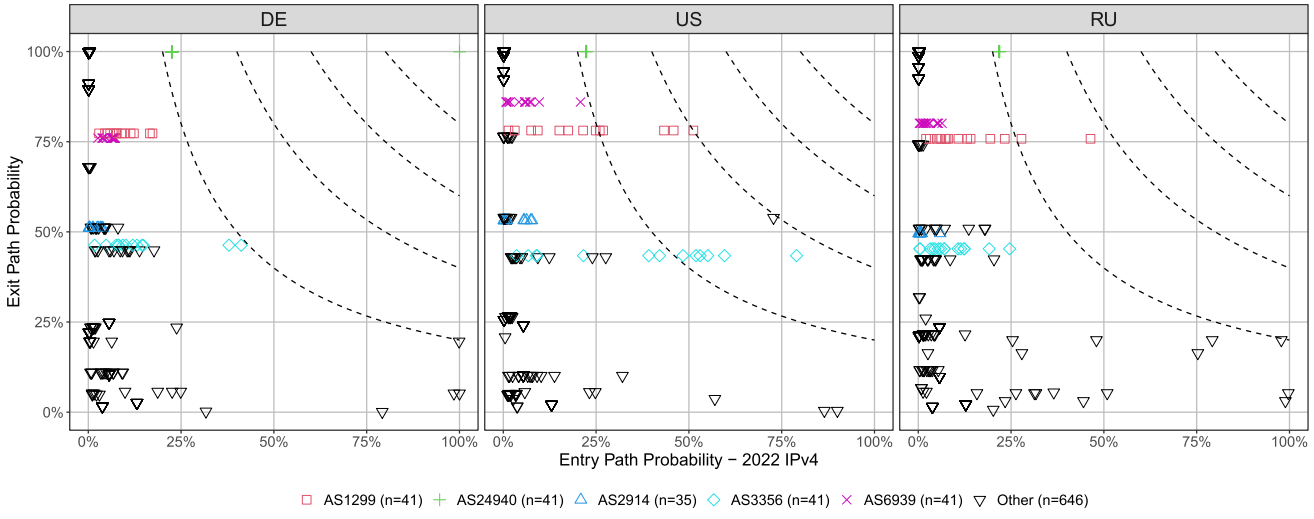
6. Discussion

Adversaries residing along the path to/from a guard relay and from/to an exit relay bear the potential to correlate traffic, thus defeating the very goal of the anonymization network Tor. In this paper, we applied our previously developed measurement methodology (Mayer et al., 2020) – capable to detect such potentially malicious players – to additional scenarios. In particular, we (a) repeated our measurements from 2020 to observe changes over time, (b) adopted our approach for IPv6 to analyze the threat when using this next-generation Internet protocol, and (c) extended our client- and destination sets to investigate the current situation in Russia where censorship intensified after its full-scale invasion of Ukraine, starting on February 24th, 2022.

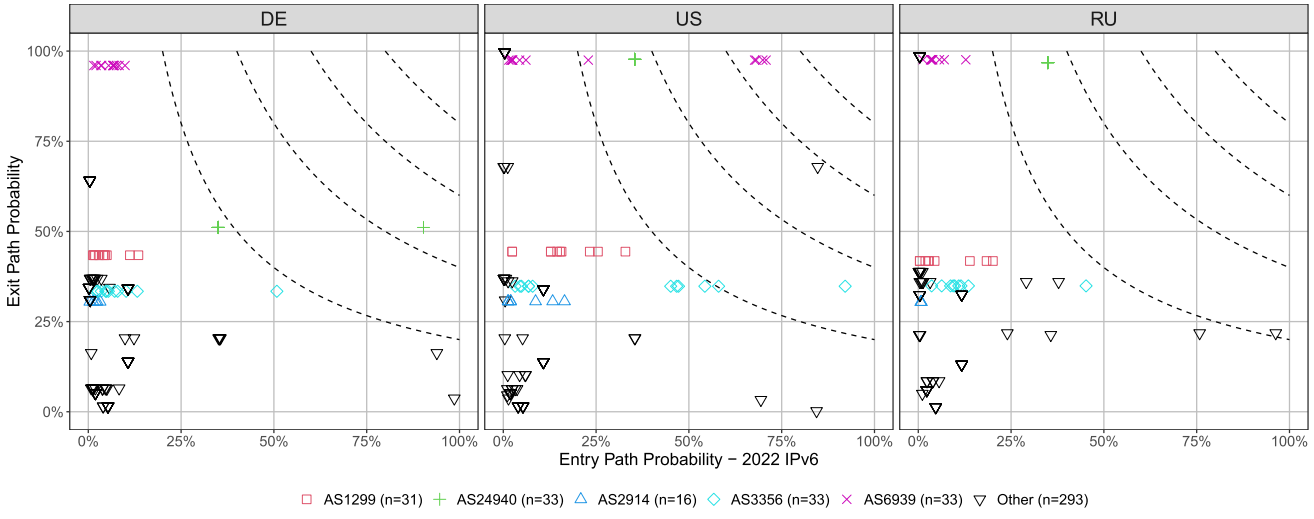
Development over time and protocols Our work does not provide any new impending AS-level adversaries. The probability of an AS to be on the entry side and/or on the exit side is – apart from a handful of changes – stable over time (2020 and 2022) and proto-



(a) 2020 IPv4



(b) 2022 IPv4



(c) 2022 IPv6

Fig. 9. ASes and their potential for traffic correlation for different years, protocols and countries for Tranco List destinations. Each data point represents an AS that appears on the entry and the exit path, and thus has the potential to perform traffic correlation. Contour lines at 20%, 40%, 60% and 80% highlight data points with highest combined probability.

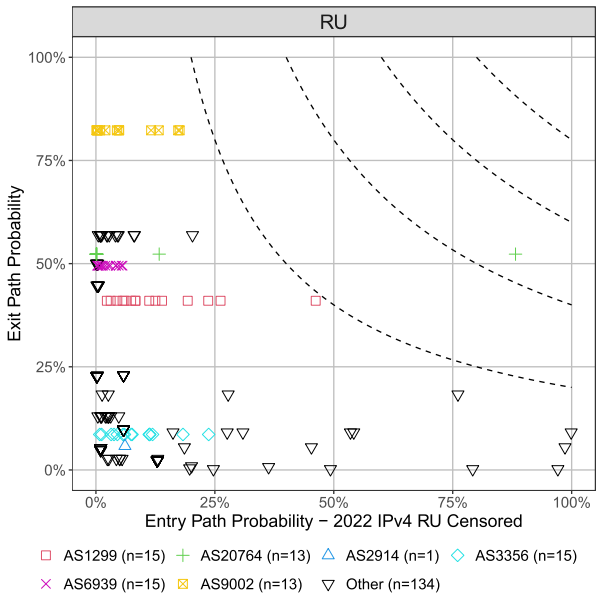


Fig. 10. ASes and their potential for traffic correlation for ASes hosting websites that are blocked by Russia. Each data point represents an AS that appears on the entry and the exit path, and thus has the potential to perform traffic correlation.

col versions (IPv4 and IPv6). This is good news: The Tor network and also the underlying routing structure of the Internet remain to a large extent stable. Tor is able to provide anonymity to users at a constantly high quality; however, targeted attacks against hand-picked combinations of clients and destinations in close proximity cannot be fully prevented (e.g., RASCOM). Beyond, it means that active measurements like ours do not necessarily have to be performed on a daily basis – longer intervals are fine, reducing the effort for measurements.

Division of roles Major transit ASes like HURRICANE or LEVEL3 are the prime suspects and are indeed capable of performing traffic correlation for many combinations of client and destination. In addition, we identified networks simultaneously serving multiple roles, which puts them in a good position for correlation attacks. For example, the data center operator HETZNER serves as a hosting provider for many destinations (e.g., major websites); at the same time, it hosts a high amount of guard relays. In total, they account for 22.4% of the guard bandwidth. This puts the AS in a favorable position for correlation attacks: The AS is likely to be part of a Tor path's entry side due to the many guard relays, and there is a high chance of it being included in the exit path due to the many hosted destinations. An operator of a HETZNER-based guard relay even found that 15% of the relay's traffic is forwarded to a relay within the same AS (Torproject, 2022a).

Ideally, guard relays should be – in network terms – close to the clients (e.g., in an ISP), and the exit guards close to the destination (e.g., in a data center), meaning that HETZNER would be a good candidate to operate exit nodes. We suggest to take this into account when deploying new guard- and/or exit relays, either as a private individual or an organization. An AS-aware circuit selection algorithm of Tor might also be beneficial but bears the risk that the chosen ASes allow to trace it back to the origin, see Section 3 on Related Work. Finally, we argue for increased AS diversity in the Tor network. Even with simple measurements, we see that the distribution of Tor relays is skewed. We hope that our measurements can improve an informed decision of how this diversity should be achieved.

Russia Since its full-scale invasion of Ukraine, Russian state authorities are blocking access to online information that is not in line with the official reports. This includes, among others, social networks, as well as local and independent media outlets. Censorship might be overcome using Tor, and our measurements show that the chance of deanonymization due to traffic correlation is low for Russian users. In fact, it is even lower than for users in Western democracies like Germany or the US (in which information censored in Russia is accessible anyway). Beyond, ASes that have the potential to perform successful correlation attacks are operated by companies in Western countries, further reducing the risk for Russian users. At the moment, however, the main challenge is to access Tor: Russian authorities aim to block guard relays, thus hindering the technology's use. The Tor community puts in a lot of effort to stay ahead of governmental blocking strategies (Dingledine, 2022).

Open source We publish our source code openly available. This enables other entities such as large relay operators to also perform measurements. All measurement results gathered through RIPE Atlas are openly available as well and could include valuable results for the Tor network. We argue that large relay operators should deploy RIPE Atlas probes in their networks, not only to further improve our (future) results but also to enable other measurements. Just a few more probes would increase the coverage significantly.

7. Limitations and future work

AS coverage Our traceroute measurements are limited by the current AS-level coverage[6] of the RIPE Atlas platform. While RIPE Atlas provides considerable coverage of a country's Internet users for Western countries (e.g., 92% in Germany and 86% in the US), its scope in illiberal or censoring states is often constrained. For example, the coverage, at the time of our measurements, was 26% in Russia, declining from 60% in 2020. Due to the current geo-politic situations and lacking alternatives, we nevertheless opted to for the inclusion of Russia as our case study. In comparison to Russia, China's Internet population is covered well by RIPE Atlas (83%), and renders it a candidate for further studies. Additionally, revisiting our measurements with increased IPv6 coverage and support among Tor relays could yield interesting results in the future.

Selection of client and destination ASes Since tracerouting all possible client and destination ASes was not feasible, we had to limit our measurements to a subset of ASes. The chosen AS sets are intended to reflect the reality best possible, i.e., the client sets should match ASes that contain actual Tor users and the destination sets destinations that are actually requested via Tor. A simple way to determine these ASes would be to capture traffic from (self-hosted) Tor relays; this, however, raises ethic concers due to snooping on Tor users and we used popular client and destination ASes instead. For our case study of Russia's full scale invasion of Ukraine, we used destinations that are blocked by the Russian regulator Roskomnadzor. We expect these destinations to be accessed via Tor as Russian Internet users cannot access them in a regular way; thus, we believe this destination set to be closer to reality than the others. Yet, there are no figures supporting this belief available.

Adversary granularity While this study specifically looks for adversaries at the granularity of ASes, there are other ways to group entities that could perform traffic correlation attacks. In some cases, organizations act as multiple ASes which means that the results (i.e., probabilities) of these ASes from our measurements have to be cumulated. Additionally, powerful nation states or intelligence agencies could force compliance of ASes within their jurisdiction to form an even more potent adversary. Finally, we executed a single traceroute for each AS pair to determine traffic

[6] https://www.sg-pub.ripe.net/petros/population_coverage/table.html

routes. Future research could provide more precise results by doing this in a more fine-grained manner, e.g., by measuring routes from different network prefixes or regions for every selected AS.

Simplified Tor model Our study is based on the traditional model of Tor covering only publicly known guard- and exit relays. In practice, Tor's architecture is constantly updated to cope with the ongoing censorship efforts of nation states like China or Russia. Therefore, Tor has introduced modular "plugable transports" (e.g., obfs4 bridges, Snowflake proxies) serving as relays which are not publicly known. This makes it harder to block these relays. We speculate that these add-ons could have positive effects on the AS distribution of the entry nodes (cf. *Division of Roles* in Section 6) due to being more lightweight, ephemeral, and easy to set up by inexperienced users (e.g., via a browser plugin). We consider this aspects to be part of future research.

8. Conclusion

We applied our measurement technology, which was developed in preliminary work (Mayer et al., 2020), to additional three use cases. This line of action allowed us to get a broader picture of current deanonymization attacks in the Tor network, and to infer actors with the potential to do so. In particular, we (a) repeated our measurements from 2020 to observe changes over time for users in Germany and the US, (b) adopted our approach for IPv6 to analyze the threat when using this next-generation Internet protocol, and (c) investigated the current situation in Russia where censorship has been intensified with the beginning of its full-scale invasion of Ukraine on February 24th, 2022.

We indeed identified a small set of ASes with the potential to perform deanonymization attacks. Most of them are large transit providers, but we also found an AS which simultaneously hosts high numbers of destinations and Tor guard relays. Hence, this AS has a high chance to appear on a Tor circuit's entry- and exit path, and consequently, successfully conducting traffic correlation to deanonymize individual Tor users. Once again, this exposes the problems of centralization and shows that there is room for improvement regarding the placement of guard-, and exit relays on the Internet. The former should be close to the clients, the latter close to the destinations.

While the numbers of individual ASes have changed since 2020, the overall picture does not reveal a significant change for Tor users in Germany and the US. Just as little does the protocol choice, i.e., IPv4 or IPv6, have a significant impact. We conclude that the Tor network provides anonymization to its users at a consistent quality. According to our results, Russian users are even less prone than Western ones to become deanonymized. Tor allows the former to securely access popular international websites as well as websites that have been censored. Beyond, the few ASes with the potential to perform successful deanonymization attacks are operated by Western companies, further reducing the risk for users in Russia.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Johanna Ullrich reports financial support was provided by Christian Doppler Research Association. Markus Maier, Florian Holzbauer, Johanna Ullrich, Georg Merzdovnik reports financial support was provided by Austrian Research Promotion Agency. Wilfried Mayer reports financial support was provided by Austrian Science Fund. Edgar Weippl is a member of the editorial board of this journal

CRediT authorship contribution statement

Gabriel K. Gegenhuber: Conceptualization, Methodology, Software, Writing – review & editing. **Markus Maier:** Data curation, Software, Validation, Visualization. **Florian Holzbauer:** Data curation, Validation, Writing – original draft. **Wilfried Mayer:** Conceptualization, Methodology, Software. **Georg Merzdovnik:** Supervision. **Edgar Weippl:** Supervision. **Johanna Ullrich:** Project administration, Funding acquisition, Writing – review & editing, Supervision.

Data availability

we have published the used source code and artifacts on github (referenced in the paper)

Acknowledgments

We want to thank David Schmidt for his preliminary work on this topic. This material is based upon work partially supported by (1) the Christian-Doppler-Laboratory for Security and Quality Improvement in the Production System Lifecycle; the financial support by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development and the Christian Doppler Research Association are gratefully acknowledged; (2) SBA Research (SBA-K1), a COMET Centre within the framework of COMET – Competence Centers for Excellent Technologies Programme and funded by BMK, BMDW, and the province of Vienna. The COMET Programme is managed by FFG; (3) Project 877110 2big2fail funded by the Program "BRIDGE1" (FFG); (4) Project DynAISEC FO999887504 funded by the Program "ICT of the Future" – an initiative of the Austrian Ministry of Climate Action, Environment, Energy, Mobility, Innovation and Technology.

Appendix A. Client and destination AS sets

A1. Client sets

$C_{2020-DE-v4}$ = {AS3320, AS6830, AS31334, AS8881, AS3209, AS6805, AS553, AS680, AS8422, AS9145}

$C_{2020-US-v4}$ = {AS7922, AS701, AS7018, AS209, AS20115, AS22773, AS5650, AS20001, AS10796, AS11427}

$C_{2022-DE-v4}$ = {AS3320, AS3209, AS8881, AS6805, AS553, AS680, AS60294, AS24940, AS8422, AS9145}

$C_{2022-US-v4}$ = {AS7922, AS7018, AS701, AS209, AS20115, AS22773, AS5650, AS20001, AS47583, AS20473}

$C_{2022-RU-v4}$ = {AS12389, AS8402, AS25513, AS42610, AS35807, AS12714, AS3216, AS8359, AS12668, AS31200}

$C_{2022-DE-v6}$ = {AS3320, AS3209, AS8881, AS6805, AS8422, AS199284, AS60294, AS24940, AS8767, AS680}

$C_{2022-US-v6}$ = {AS7922, AS7018, AS701, AS47583, AS20473, AS62538, AS20001, AS209, AS22773, AS20115}

$C_{2022-RU-v6}$ = {AS42610, AS25513, AS202422, AS8331, AS12668, AS20764, AS50716, AS35807, AS12714, AS15974}

A2. Destination sets

$D_{2020-v4}$ = {AS3, AS15169, AS4837, AS24940, AS36351, AS14618, AS16509, AS14907, AS3356, AS7941}

$D_{2022-TRANCO-v4}$ = {AS15169, AS16509, AS8075, AS4837, AS14907, AS55990, AS37963, AS132203, AS4134, AS4812, AS47764, AS29169, AS14618, AS396982}

$D_{2022-TRANCO-v6}$ = {AS15169, AS16509, AS14907, AS47764, AS63949, AS3, AS37963, AS197695, AS32, AS14618}

$D_{2022-RU-CENSORED-v4}$ = {AS200350, AS15497, AS25532, AS207651, AS9123, AS28907, AS3326, AS197695, AS25521, AS12722}

References

Akhoondi, M., Yu, C., Madhyastha, H.V., 2012. LASTor: a low-latency AS-aware Tor client. Symposium on Security and Privacy. IEEE.

Alsabah, M., Goldberg, I., 2016. Performance and security improvements for Tor: A survey. ACM Comput. Surv. (CSUR) 49 (2), 1–36.

Barton, A., Wright, M., 2016. DeNASA: destination-naive AS-awareness in anonymous communications. Proc. Privacy Enhanc. Technol. 2016 (4).

Barton, A., Wright, M., Ming, J., Imani, M., 2018. Towards predicting efficient and anonymous Tor circuits. USENIX Security Symposium.

Dingledine, R., 2013. The lifecycle of a new relay. https://www.blog.torproject.org/lifecycle-new-relay.

Dingledine, R., 2022. How Russia is trying to block Tor. https://www.media.defcon.org/DEFCON30/DEFCON30presentations/RogerDingledine-HowRussiaistryingtoblockTor.pdf.

Dingledine, R., Mathewson, N., Syverson, P., 2004. Tor: The Second-Generation Onion Router:. Technical Report. Defense Technical Information Center, Fort Belvoir, VA doi:10.21236/ADA465464. http://www.dtic.mil/docs/citations/ADA465464

Eaton, E., Sasy, S., Goldberg, I., 2022. Improving the privacy of Tor onion services. In: Ateniese, G., Venturi, D. (Eds.), Applied Cryptography and Network Security. In: Lecture Notes in Computer Science, vol. 13269. Springer International Publishing, Cham, pp. 273–292. doi:10.1007/978-3-031-09234-3_14.

Edman, M., Syverson, P., 2009. AS-awareness in Tor path selection. In: Conference on Computer and Communications Security. ACM.

Feamster, N., Dingledine, R., 2004. Location diversity in anonymity networks. Workshop on Privacy in the Electronic Society. ACM.

Gill, P., Schapira, M., Goldberg, S., 2012. Modeling on quicksand: dealing with the scarcity of ground truth in interdomain routing data. ACM SIGCOMM Comput. Commun. Rev. 42 (1), 40–46.

Giotsas, V., Luckie, M., Huffaker, B., kc claffy, 2014. Inferring complex AS relationships. In: Internet Measurement Conference. ACM.

Hopper, N., Vasserman, E.Y., Chan-Tin, E., 2010. How much anonymity does network latency leak? ACM Trans. Inf. Syst. Secur. (TISSEC) 13 (2), 1–28.

Jansen, R., Johnson, A., 2021. On the accuracy of Tor bandwidth estimation. In: Hohlfeld, O., Lutu, A., Levin, D. (Eds.), Passive and Active Measurement. In: Lecture Notes in Computer Science, vol. 12671. Springer International Publishing, Cham, pp. 481–498. doi:10.1007/978-3-030-72582-2_28.

Johnson, A., Wacek, C., Jansen, R., Sherr, M., Syverson, P., 2013. Users get routed: traffic correlation on Tor by realistic adversaries. In: Conference on Computer and Communications Security. ACM.

Juen, J., Johnson, A., Das, A., Borisov, N., Caesar, M., 2015. Defending Tor from network adversaries: a case study of network path prediction. Proc. Privacy Enhanc. Technol. 2015 (2), 171–187.

Mayer, W., Merzdovnik, G., Weippl, E., 2020. Actively probing routes for Tor as-level adversaries with ripe atlas. In: IFIP International Conference on ICT Systems Security and Privacy Protection. Springer, pp. 234–247.

Hanley, H., Sun, Y., Wagh, S., Mittal, P., 2019. DPSelect: a differential privacy based guard relay selection algorithm for Tor. Proc. Privacy Enhanc. Technol., 2019 (2).

metrics, T., 2022. Servers. Retrieved Sept. 19, 2022 from https://www.metrics.torproject.org/bubbles.html/as-exits-only.

Migliano, S., Woodhams, S., 2022. Websites blocked in russia since ukraine invasion. Retrieved Sept. 30, 2022 from https://www.top10vpn.com/research/websites-blocked-in-russia/.

Mittal, P., Khurshid, A., Juen, J., Caesar, M., Borisov, N., 2011. Stealthy traffic analysis of low-latency anonymous communication using throughput fingerprinting. In: Conference on Computer and Communications Security. ACM.

Nasr, M., Bahramali, A., Houmansadr, A., 2018. DeepCorr: strong flow correlation attacks on Tor using deep learning. In: Conference on Computer and Communications Security. ACM.

Nithyanand, R., Starov, O., Zair, A., Gill, P., Schapira, M., 2016. Measuring and mitigating AS-level adversaries against Tor. Network and Distributed System Security Symposium (NDSS).

Pochat, V.L., Goethem, T.V., Tajalizadehkhoob, S., Korczynski, M., Joosen, W., 2019. Tranco: a research-oriented top sites ranking hardened against manipulation. Network and Distributed System Security Symposium.

Rochet, F., Wails, R., Johnson, A., Mittal, P., Pereira, O., 2020. CLAPS: client-location-aware path selection in Tor. In: Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. ACM, Virtual Event USA, pp. 17–34. doi:10.1145/3372297.3417279.

Staff, R., 2015. RIPE atlas: a global internet measurement network. Internet Protoc. J. 18 (3), 2–26.

Sun, Y., Edmundson, A., Feamster, N., Chiang, M., Mittal, P., 2017. Counter-RAPTOR: safeguarding Tor against active routing attacks. Symposium on Security and Privacy. IEEE.

Sun, Y., Edmundson, A., Vanbever, L., Li, O., Rexford, J., Chiang, M., Mittal, P., 2015. RAPTOR: routing attacks on privacy in Tor. USENIX Security Symposium.

Torproject, 2021. The state of IPv6 support on the Tor network. Retrieved Sept. 30, 2022 from https://www.blog.torproject.org/state-of-ipv6-support-tor-network.

Torproject, 2022a. Are Hetzner servers in both the guard and middle position for a lot of Tor circuits? Observations from Hetzner traffic numbers vs. own monitoring. Retrieved Sept. 30, 2022 from https://www.forum.torproject.net/t/are-hetzner-servers-in-both-the-guard-and-middle-position-for-a-lot-of-tor-circuits-observations-from-hetzner-traffic-numbers-vs-own-monitoring/1851.

Torproject, 2022b. Tor protocol specifications. Retrieved Sept. 14, 2022 from https://www.gitweb.torproject.org/torspec.git/tree/.

Vanbever, L., Li, O., Rexford, J., Mittal, P., 2014. Anonymity on QuickSand: using BGP to compromise Tor. In: Proceedings of the 13th ACM Workshop on Hot Topics in Networks. ACM.

Wacek, C., Tan, H., Bauer, K.S., Sherr, M., 2013. An empirical evaluation of relay selection in Tor. Network and Distributed System Security Symposium.

Wan, G., Johnson, A., Wails, R., Wagh, S., Mittal, P., 2019. Guard placement attacks on path selection algorithms for Tor. Proc. Privacy Enhanc. Technol., 2019(4).

Xynou, M., Filastò, A., 2022a. New blocks emerge in russia amid war in ukraine: an OONI network measurement analysis. Retrieved Sept. 30, 2022 from https://www.ooni.org/post/2022-russia-blocks-amid-ru-ua-conflict/.

Xynou, M., Filastò, A., 2022b. Russia started blocking Tor. Retrieved Sept. 30, 2022 from https://www.ooni.org/post/2021-russia-blocks-tor/.

Gabriel K. Gegenhuber is a Ph.D. student at University of Vienna, Austria, Research Group Security and Privacy. Gabriel received a B.Sc. in Software & Information Engineering and an M.Sc. in Software Engineering & Internet Computing at the Technical University of Vienna. His research interests include mobile networks, network measurements, network security, and privacy-enhancing technologies.

Markus Maier is a Ph.D. student at University of Vienna, Austria, Research Group Security and Privacy. He received his B.Sc. and M.Sc. at Technical University of Vienna in Software Engineering & Internet Computing. His research interests include routing, network measurement and network security.

Florian Holzbauer is a Ph.D. student at University of Vienna, Austria, Research Group Security and Privacy. He received his B.Sc. and M.Sc. at University of Applied Sciences St.Pölten. His research focuses on Internet measurements. He detected flaws in email-related protocol adoption and is currently looking for flaws in IPv6 deployments.

Wilfried Mayer received a master's degree in Software Engineering and Internet Computing, and a doctoral degree in computer science at TU Wien. His research interests are focused on measuring privacy-enhancing technologies.

Georg Merzdovnik received a B.Sc. in computer engineering, an M.Sc. in software and information engineering, and a Ph.D. in computer science with distinction at TU Wien. Currently, he leads the research group on Systems and (I)Iot Security at SBA Research. Georg's research interests include applied systems and software security, IoT security (ranging from device to network level) as well as online privacy in general.

Edgar Weippl Edgar graduated with a Ph.D. from TU Wien. Afterwards, he was an assistant professor at Beloit College, WI, and a consultant for the software vendor ISIS Papyrus in New York, NY, Albany, NY, and Frankfurt, Germany. Returning to Vienna, he co-founded the research center SBA Research in 2004. In 2020, Edgar accepted a position as full professor at the University of Vienna.

Johanna Ullrich received a Ph.D. sub auspiciis praesidentis from TU Wien. Currently, she is a key researcher at SBA Research, Austria, leading the Networks and Critical Infrastructures Security Research Group, and a researcher of the Christian Doppler laboratory SQI. She was awarded the Research Prize of the Dr. Maria Schaumayer Foundation and nominated for the Hedy Lamarr Prize twice. Her research focuses on network security, particularly measuring experiments and IPv6.